Chapter 4

On the phonetic basis of phonological change

Patricia Donegan

1 The nature of phonological change

1.1 Sound changes are changes in speakers' phonetic abilities
The distinction in N. American English between the low labial vowel [o] of caught, dawn, hawk and the low non-labial vowel [a] of cot, Don, hock is an old one, going back to early English aw versus short o. But in dialects of Canada and various parts of the United States of America, the [o] is pronounced [a], and the two sets of words are no longer distinguished. This sound change [o] > [a] was a historical event, but it remains a synchronic constraint that characterizes the phonetic capacities of the speakers of the innovating dialect: without additional learning, they cannot perceive or pronounce the [o] of other speakers, and even if they learn to perceive it, they often continue to pronounce it as [a] (at least in unguarded moments) or as some other substitute, e.g. [o] or [o] or [o]. It is significant that children learning the conservative dialect also typically at first cannot perceive and pronounce [o]. When they learn to perceive it, they still use [a] or [o] or [o] or [o] until they learn also to pronounce it. Thus speakers of the conservative dialect acquire a phonetic ability that speakers of the innovating dialect do not.

But some adult speakers of dialects with [o] in fact never learn to perceive it or to pronounce it without substitution. They are innovators, though their innovation lies in not doing something their fellow speakers do do, namely learn to perceive and pronounce [o]. And some speakers, having learned to hear and say [o] as children, may move to an innovative dialect area and, under peer influence, lose the ability to even to hear [o] as distinct from [a]. One such person, whose childhood distinction of [o] and [a] was well attested in a linguist's diary, later forgot
ever having distinguished caught from cot himself, and even refused to believe that anyone else could do so. After hearing it demonstrated by a perfect commutation test, he suspected a trick, and demanded a replay. At last he said, 'I get it: you're saying [koqt]!' (for caught). But in fact the subjects had said only [kot] for caught. The sceptic, having finally heard [o], could reproduce it only as the substitute [og].

1.2 Phonological processes respond to universal phonetic difficulties

The phonetic qualities that cause a vowel like [o] to be difficult to perceive as such, and also difficult to pronounce as such, are rather simple. [o] is a maximally open vowel produced with the lips narrowed by rounding. Opening the mouth wide conflicts with rounding the lips. The conflict attenuates one or the other acoustic impression, and it also makes it hard to produce both acoustic impressions simultaneously. The difficulty is obviously universal, and so are the possible remedies: either to give up the low or the labial feature, or to sequence them in a diphthong. That is,

(1) Low vowels must be non-labial, i.e. [o] \(\rightarrow\) [a]
(2) Labial vowels must be non-low, i.e. [o] \(\rightarrow\) [o].
(1+2) [o] \(\rightarrow\) [oQ]
(2+1) [o] \(\rightarrow\) [oQ]

Besides being heard in children learning [o] in English and other languages, substitution (1) is standard in the North American dialects already mentioned, (2) in British Received Pronunciation (RP) (Gimson 1962: 109ff.), (1+2) in some southern United States dialects (cf. Hall 1942, Kurath and McDavid 1961), and (2+1) in northern United States urban areas (Labov et al., 1972). Since the sequenced articulation of diphthongs takes more time, often the diphthongizations are limited just to long or contextually lengthened vowels.

Since [o] shares with [o] the conflict of rounding and openness, but has only medium opening, we would predict parallel but less frequent substitutions for [o], and all of them in fact occur: (1) IE short \(\ast o >\) Sanskrit [\(\lambda\)] (written a, e.g. dāma- 'house'); (2) OE long [o:] \(>\) late ME [u:], e.g. fool; (1+2) British RP and Middle Atlantic [o] \(\rightarrow\) [oQ] and often [oQ] \(\rightarrow\) [εo], e.g. snow; and (2+1) Germanic long \(\ast o >\) Old High German [uο] (spelled uo, e.g. buoch 'book', but still pronounced [uο] in conservative German dialects). As the hierarchy of difficulty would suggest, none of these substitutions affects [o] unless it also affects [o].
if [o] is raised to [u], [o] is raised to [o].

There are parallel substitutions for the palatal vowels [æ] and [e]. Since it is difficult to combine a palatal tongue position with an open jaw, one can be given up:

(3) *Non-high vowels must be non-palatal*, i.e.

\[\text{[æ]} \rightarrow \text{[o]}, \text{e.g. British and Northeastern US can't, past; }\]
\[\text{[e]} \rightarrow \text{[ɛ]}, \text{e.g. New Zealand dialects let, said};\]

or

(4) *Palatal vowels must be non-low*, i.e.

\[\text{[æ]} \rightarrow \text{[e]}, \text{e.g. English Vowel Shift meat; }\]
\[\text{[e]} \rightarrow \text{[i]}, \text{e.g. English Vowel Shift meet; and again in meat.}\]

Or the conflicting features can be sequenced:

(3+4) \[\text{[æ]} \rightarrow \text{[o:]}, \text{e.g. Southern US dialects bad; }\]
\[\text{[e]} \rightarrow \text{[aj]}, \text{e.g. Australia, NZ, North Carolina bay.}\]

(4+3) \[\text{[æ]} \rightarrow \text{[e]}, \text{e.g. Northern urban US dialects bad; }\]
\[\text{[e]} \rightarrow \text{[i]}, \text{e.g. Old High German hiar 'here'.}\]

The low round palatal vowel [æ] combines the difficulties of [o] with those of [æ], and adds the conflicting acoustic effects of tongue fronting and lip rounding. Such vowels are rare, learned late, and historically, lost early. Even the easiest of the labiopalatals, the high front rounded vowel [y], is usually unrounded (e.g. OE *my:s* > ME *mǐs* 'mice'), or diphthongized (e.g. MHG *myːz:* > NHG *[moiz~]* 'id.').

But even the optimal palatal vowel [i] and the optimal labial vowel [u] present difficulties. For children, even simple palatalization or labialization can be difficult, and since it is children who learn languages, it should not be surprising that there are languages like Adyghe, Kabardian, and others of the northwest Caucasus (Catford 1977), Gude (Hoskison 1974), or Marshallese (Bender 1971) in which the only high vowel phoneme is non-palatal and non-labial [i] – i.e., in which both depalatalization and delabialization apply. Furthermore, since the optimal vowel would be maximally audible, like [o], high vowels are not optimal. So we find languages in which they are lowered; e.g. Squamish, Alabama, and Amuesha lack high vowels (Maddieson 1984). And in English and many other languages, we find diphthongizations of [i] and [u] that lower their nuclei to [ɛi], [ɔi]; or depalatalize or delabialize them to [i], [ui]; or that do both, to [ɛi], [ɔi], and lower again to [qi], [ou]. Of course, this was the path of early English long [i:] and [u:] as in *mice* and *mouse*.
We are left with what Hellwag called the Prince of Vowels, \( \text{[a]} \), which would seem to present no difficulties at all: it is optimally open and has no conflicting features like palatality or labiality. If a child can say only one vowel, that vowel is \( \text{[a]} \). Or if a language has only one vowel, we would expect it to be \( \text{[a]} \). And we might expect that \( \text{[a]} \) would always remain \( \text{[a]} \). But we find that \( \text{[a]} \) sometimes does undergo substitution, acquiring either palatality \( ([\text{a}] \rightarrow ([\text{æ}])) \) or labiality \( ([\text{a}] \rightarrow ([\text{o}])) \), or losing some of its sonority \( ([\text{a}] \rightarrow ([\text{ʌ}])) \). These changes frequently occur in languages where short and long \( \text{[a]} \) are distinguished lexically, but where the spoken length is disturbed by prosodic changes like stress timing, which lengthens and shortens vowels to fit their rhythmic context. A common response to this is the superimposition of a tense-lax distinction on the long-short distinction. But since tenseness in vowels is an intensification of palatal or labial colouring (Donegan 1978), it is inapplicable to vowels like \( \text{[a]} \), which are neither palatal nor labial. For example, Latin long vs. short vowels became proto-Romance tense vs. lax vowels, but Latin long and short \( \text{[a]} \) merged. The imposition of palatal or labial colour on long or short \( \text{[a]} \) – or on both, so that the tense-lax difference would be possible – could reinforce a threatened length distinction and prevent lexical mergers. For example, OE long \( \text{[a]} \), in words like \text{stan} ‘stone’ and \text{râp} ‘rope’, became \( ([\text{a}]:) \) (long open \( \text{[a]} \)) in early ME, while short \( \text{[a]} \) remained \( \text{[a]} \). Later, in Middle English, long \( \text{[a]} \) (from lengthened \( \text{[a]} \), as in \text{mate} ) was fronted to \( \text{[æ]} \), while short \( \text{[a]} \) remained \( \text{[a]} \). Hock (1986: 144) cites Modern Persian as a language where both labialization and palatalization have applied, one to long \( \text{[a]} \) \( \rightarrow ([\text{a}:]) > ([\text{o}]) \), and the other to short \( \text{[a]} \) \( ([\text{a}] > ([\text{æ}]) \).

Since the disadvantage of combining high sonority with either palatal or labial colouring remains, we should expect \( \text{[a]} \) to be less susceptible to such colouring than higher vowels are. And we do find that, other things being equal, \( ([\text{a}] \rightarrow ([\text{o}])) \) implies that, in the same speakers, \( ([\text{ʌ}] \) in the system \( \rightarrow ([\text{ʊ}]) \). Similarly, \( ([\text{ʌ}] \rightarrow ([\text{ʊ}]) \) implies \( ([\text{i}] \rightarrow ([\text{ʊ}]) \) (Donegan 1978). Cross-language comparison of changes reveals, then, that opposite substitutions have opposite implicational conditions on their application – e.g. the more sonorant a vowel is, the more susceptible it is to delabialization, but the less sonorant a vowel is, the more susceptible it is to labialization. Far from making us conclude that ‘anything can happen’ and there are no phonetic motivations, careful comparison of changes confirms – and indeed often reveals – their phonetic motivations.
1.3 Phonological processes are not physical slips
It is important to realize that these systematic substitutions that result in phonetic change do not occur by 'accident'. The vocal tract itself cannot select a substitution to resolve a phonetic difficulty; it must be the brain, or the mind, that chooses a substitute that is both regular and perceptually similar to its input. The substitutions that occur are motivated by vocal and perceptual constraints, but they apply in the mind. It is the mind that gives the instructions that produce a substitution. It is likewise the mind – not the ear or the mouth – that learns to perceive and pronounce the difficult segment or sequence. Only the mind can attend (or not attend) to the appropriate cues in perceiving sounds as the same or different, and only the mind can consistently give the appropriate neuromuscular instructions to produce the 'output' of the substitution. The processes are mental, because it is the mind that, having insufficient control of the body, determines or discovers a regular substitute. And it is the mind that, in learning to control articulation and perception, learns the abilities or 'keys' that turn off the substitutions.²

But the mental nature of processes does not require us to assume that phonological processes are learned through the observation and comparison of forms, as morphology must be learned. Unlike morphology and morphophonology, which are conventional, phonological processes are motivated by the demands of the body, not the results of analysis.

2 The phonological system as the representation of phonetic abilities: phonetic change is phonological
The set of phonetic constraints and responding substitutions that speakers are born with can be represented as a set of phonological processes. Each process replaces a class of sounds which presents a difficulty to the speaker with another class which lacks that particular difficulty. These processes represent the speaker's inabilities. What the speaker learns, in learning the phonology of a particular language, is the required set of abilities to perceive and pronounce. That is, in learning to perceive and pronounce the required segments and sequences, he acquires the abilities or 'keys' to turn off the processes which would eliminate these sounds or sequences. We can represent these learned abilities as inhibitions of phonological processes. The set of constraints and substitutions – phonological processes – is universal. What a linguistic community shares is the set of learned inhibitions of these processes.
In our first example, the conservative speakers who distinguish /ɒ/ from /a/ have learned to inhibit the process represented by (1). When a 'phonetic change' takes place, speakers fail to learn the 'key' that turns off the process that is involved. In the case of an optional or variable change, speakers learn to inhibit the process and pronounce its input, but their learning is imperfect. Then using this ability or key requires special attention or effort, so they may fail to do so in certain prosodic or stylistic circumstances.3

Any difficulty that is not mastered is avoided automatically by substitution. The phonetically motivated, universal processes which the speaker continues to apply (either obligatorily, or as options) after he has acquired the adult form of the language are responsible for the phonological substitutions of the language. If, through some failure of learning, or through some relaxation of the requirements of the community (even if only in limited contexts), a speaker continues to make a substitution that was not generally accepted at an earlier time, he may appear to have 'added' a phonological process to the language. Such substitutions, when context-sensitive, may create variant pronunciations without changing the phonological form, because the speaker can attribute the variants to their occurrence in a particular environment. For example, the de-rhotacization of syllable-offset r, in Eastern New England and Metropolitan New York ear, care, door [ɪə, keə, dɔə] (Kurath and McDavid 1961: 171), does not change their phonemic status as /ɪr, ker, dɔər/, where they are pronounced with [r] before a vowel. But this 'added' r-deletion also does not represent an invention, or the creation of some new linguistic operation. The innovator has not 'made a change' in his own linguistic system. In a sense, he has done nothing. He has simply continued to submit to a constraint that other speakers overcome, because he has failed to acquire the ability to inhibit the resulting substitution.

2.1 Phonology is not conventional
Unlike the 'natural' theory being described here, most theories of phonology, structuralist and generative, are conventionalist theories. They regard phonological substitutions as grammatical conventions that are discovered, or learned, by speakers on the basis of their analysis of language data. These conventionalist theories — generative phonology, in particular, and its autosegmental, metrical, and lexical variants — do not distinguish in kind between phonological processes, which are based entirely on phonetic and prosodic features, and morphophonological rules.4
Morphophonological rules are purely conventional, or traditional. They may resemble phonological processes because they can be expressed, in part, in phonological features. But they apply only in connection with grammatical processes and in terms of lexical and grammatical categories, not prosodic ones. And they apply only in terms of phonemes, or classes of phonemes that may be phonetically arbitrary (e.g. in English, velar softening relates 'hard' /k, g/ to 'soft' /s, j/).

This means that morphophonological rules do not represent constraints on pronunciation. That is, their inputs are not difficult or unpronounceable. In English, for example, there are obligatory morphophonological rules that make k → s in words like electric + ity [ɪlek'trisiti], and that make t → null in words like chaste + en [ˈɛɪʃtn]. But any English speaker who can pronounce persnickety [ˈprəsnɪkiti] or Boston [ˈbostən] could perfectly well pronounce electric + ity as [ɪlek'trɪkiti], or chaste + en as [ˈɛɪʃtn]. Rules like these may have their historical origins in alternations of phonemes that were originally phonetically motivated, but they no longer represent limitations on what the speaker finds pronounceable. Instead, they have become conventions or traditions to which the speaker conforms in order to make his speech grammatical. (For further discussion and examples, cf. Darden 1989.)

The absence of this distinction between phonetically motivated and conventional substitutions – and the consequent assumption that phonological substitutions are conventional – create a gap between phonetic change and phonology that is virtually impossible to reconcile. Scholars like Labov, Ohala, Janson, and others have documented phonetic change and explored the reasons for such change, but they have not explained how the phonetic changes are implemented in speakers’ phonological systems, or how they affect these systems. For example, Labov, Yaeger, and Steiner, in their 1972 vowel shift study, documented a remarkable number of vowel changes in many dialects of English, compared these changes to vowel shifts in the histories of other languages, and attempted to specify generative rules by which these changes were taking place. But the status of these rules in the innovating speakers’ phonological systems, the question of why or how speakers would adopt such rules, and the ways in which the rules affected the speakers’ phonological representations were not addressed.

It is sometimes assumed that changes are introduced ‘in the phonetics’ of the innovating speakers, and that other speakers invent a rule in imitation of the innovators. But this leaves in
question the nature of the change in the innovators' phonetic processing.

Conventionalist theory, where phonological rules (like morphophonological rules) must be learned, fails to explain systematically how phonetic motivation is related to the adoption of new phonetic targets, how phonetic rules are acquired, or what it means for these new rules and targets to become phonological.

2.2 The identity of phonology and phonetics

While conventionalist phonology has obscured the difference between phonology and morphophonology, it has seemed to exaggerate the distinction between phonology and phonetics, and to assign much of the detail of pronunciation and change to the phonetic component. But phonology is phonetic processing; it is thus the mental (or central-nervous-system) aspect of phonetics. Although certain phonetic events may occur in the vocal tract of the speaker, or in the air, or in the ear of the hearer, phonetic processing is a matter of the perception and the production of sounds, and it is thus fundamentally mental. Indeed, when we speak and hear in our imaginations, even though no physical articulations or sounds actually occur, the constraints remain the same as they are in actual speech. The nature of articulations and the acoustic or perceptual character of sounds, of course, affect speakers' categorizations. Such physical factors also motivate the alterations or substitutions speakers make. But the categorizations and alterations themselves are mental. The relationship between the physical events and the mental processes is what makes phonetics different from physiology or acoustics.

Thus, phonology must include matters of 'phonetic detail'. Unless we are willing to believe that differences between sounds in different languages or dialects are the result of differences in their speakers' vocal tracts, we must regard such differences as results of the processing of speech. Articulatory parameters such as speech rate and jaw mobility, or the particular gestures used to implement features like voicing or implosion in obstruents or frontness (palatality) in vowels may vary from speaker to speaker. Such differences may be idiosyncratic and vary randomly from speaker to speaker, making little difference in the perceptual quality, in which case they are irrelevant to phonology. But if such differences turn out to be consistent across speakers within particular communities, and particularly if they are associated with differences in the applications of substitutions, then they are part of the phonology.

Correspondingly, a systematic innovation in a speaker's
pronunciation, even if it is only a small difference of articulatory target, or a difference in the timing of articulatory gestures, must result from a difference in the processing of speech. It cannot be accomplished by the vocal tract itself, but must result from some specific relaxation of the requirements of the language in favour of some requirement (whether fortitive or lenitive) of the speaker. If phonology were a matter of conventional or merely traditional substitutions, we would have to explain how substitutions become 'phonologized', or 'grammaticized'. But if phonology is phonetic processing, through a set of systematic substitutions with direct phonetic motivation, then any phonetic change occurs in the phonology, by the application of a substitution that other speakers inhibit.

2.3 The identity of phonological and phonetic features
If phonological processes specify phonetic detail, and if the conditions on such processes are sensitive to phonetic detail, such detail must be reflected in the system of phonological features. And, since phonological processes have both articulatory and perceptual motivations, the features in terms of which they operate must integrate these. One can conclude that phonological features are the mental connections speakers draw between particular gestures or commands to the vocal organs and the perceived effects of those gestures or commands. There is no distinction, in this view, between phonetic features and phonological features.

Phonological features, then, are universal. Segments specified with the same features ought to be the same from language to language, articulatorily and acoustically (allowing, of course, for individual variation). But segments said to have the same features often are not the same from language to language. For example, English initial voiceless stops are said to be aspirated. Yet speakers of Hindi and other Indian languages with aspirated voiceless consonants never borrow English initials as aspirates. This simply suggests that features currently in use do not include sufficient detail, and are thus inadequate to account for phonological substitutions (or for phonological change). This is not surprising, considering their origins in the minimal feature set sufficient to specify only the phonemic oppositions occurring in languages (Jakobson et al. 1963). Feature sets in general use typically continue to specify 'contrasts observable at the systematic phonetic level', even though 'there is no clear evidence showing that the set of features required for specifying phonetic contrasts is the same as that required for specifying the natural classes of
sounds required in phonological rules' (Ladefoged 1971: 1).

Distinctions could actually be represented by speakers in terms of sounds. It is only because speakers treat similar sounds identically with respect to particular substitutions that we can claim that features are part of the actual processing of speech. That is, the only real evidence for the psychological or linguistic reality of phonological features is their function in synchronic or diachronic phonological substitutions (cf. Keating 1988). Substitutions may specify features that are non-distinctive, and they are often, in turn, conditioned by features which are non-distinctive. For example, in some of the southern United States dialects noted earlier (s. 1.2), /æ/ and /ə/ diphthongize to [æe] and [oʊ] only when lengthened (before voiced stops or fricatives, in monosyllables), even though length is not distinctive. In fact, features which are not distinctive in any language (e.g. release of stops) may affect the application of processes. Release, for example, blocks assimilation of an apical stop to a following non-apical in English (bad cat [bæg kʰæt] but not *[bæg kʰæt]). It also blocks assimilation (deletion) of stops after fricatives or nasals in words like fast, or hand. Thus a non-distinctive feature controls processes that neutralize feature values that are distinctive.

It would seem, then, that feature theory must concern itself with substitutions and their conditioning factors and with the relationship between the features relevant in processing and their phonetic correlates. Ladefoged (1980) has identified some seventeen articulatory and fifteen acoustic parameters that can specify the actual sounds of languages, but he notes that these parameters are not in a direct or one-to-one relationship to phonological features. There is, in fact, no reason to assume that the relationship between phonetic parameters and phonological features would be simple, if features are speakers' ways of relating articulation to perception. But until the relationships between phonological features and articulatory and acoustic parameters can be established, phonological features will be inadequate to account for phonological substitutions or change.

2.4 The scope of phonology
In the view I am expounding, phonology is a system of processes embodying the entire faculty for perceiving and pronouncing speech (or misperceiving and mispronouncing it), plus those 'keys' – inhibitions of the processes – that the speaker acquires in learning to pronounce a language. Strictly speaking, the inhibitions would be sufficient as a description of what is language-particular, and a more parsimonious kind of phonological
description is hard to imagine. But the point of recognizing the identity of phonology and phonetics is not to eliminate the universal or the predictable from language descriptions. Our aim is not parsimony; it is understanding, and clearly, we can understand phonology — the discrepancy between what we say and what we think we are saying — only if we understand the processes that cause that discrepancy.

The focus of structuralist and generative phonology on rigorous justification of descriptions of languages has eliminated or marginalized anything 'external' (child speech, baby talk, foreignisms, singing, verse, etc.) and has limited discussion of anything not 'distinctive' or 'standard' (allophonics, individual and dialectal variation, etc.) Many 'phonetic realization rules', or 'default specifications' are omitted or consigned to a separate phonetic component. This is understandable: what possible relevance can the 'default' phonetic process that raises low labial vowels have to a description of Japanese, for example, which has no low labial vowels? But the result is phonological descriptions that end just when they are getting interesting, ignoring the details of rhythm and phonetic nuance that are most characteristic of the truly unique 'accent' of its speakers. And they stop short of the very questions that that unique 'accent' poses, e.g. why Japanese speakers who learn English words should perceive and pronounce [b] as [o].

Under the influence of this focus on the minimal relevant characteristics of speech, the role of phonetics in phonology was reduced to providing descriptions of phonemic distinctions or distinctive features. But it is clear that phonemes are not the sums of their distinctive features, and that phonological processes do not operate only in terms of phonemes and distinctive features. Before structuralism, the standard handbooks of phonetics covered such 'non-distinctive' features as syllabication, and also the phonological processes that are responsible for free and contextual variation. Whether we call it phonetics or phonology, all this is essential to understanding how speakers of any language perceive and produce speech, and how they are prepared for situations that could never have been anticipated in language learning: having to pronounce foreign words, or the results of tongue slips or Pig Latins, or having to cope with extraordinary tempos, or noise, or fatigue, or a dentist's tools in one's mouth. Obviously, no minimalist description of a language will ever tell us what pronunciations the language could have in the next generation. But phonology should.
3 The phonetic basis of phonemic systems

3.1 Fortitions and lenitions
The phoneme inventory and the phonemic forms of a language arise, in each learner, in the interaction between two sets of phonetic demands (cf. Stampe 1987). One set of demands works to limit the inventory of 'possible' – intentionally pronounceable and perceivable – segments in a language; these demands are represented by fortition processes. Fortitions are typically context-free or dissimilative, and they apply to optimize individual segments in some way; they overcome difficulties that are associated with simultaneous combinations of conflicting features, or they optimize a feature of the individual segment. Delabialization and depalatalization, as exemplified in (1) and (3), and raising, as exemplified in (2) and (4), are examples of fortitions.

The other set of demands on the speaker works to make the 'possible' segments pronounceable in the different contexts in which they appear: these are represented by lenitions. Lenitions apply to optimize sequences of segments; they overcome difficulties associated with sequential combinations of features. Assimilative processes, like vowel nasalization before nasals and consonant palatalization before palatal vowels, are examples of lenitions.

It is in the speaker's phonetic interest to have as few distinct segments as possible to produce or perceive or remember, and to allow fortitive processes to constrain the possibilities. Therefore, in languages like Japanese or Spanish, with five-vowel systems /i/, e, a, o, u/, fortitive processes – like unrounding (1) or depalatalization (3), or raising (2) or (4) – eliminate /æ/ and /ə/. These processes are part of the speaker's native incapacity, since they represent things the speaker has not learned and does not have to learn. They represent the inability to pronounce or perceive their inputs (low palatal or low labial vowels), and the requirement that the speaker substitute their outputs.

In applying, fortitions limit the possible segment inventory, so the speaker encodes forms lexically only in terms of a limited set of relatively optimal 'possible' segments. Obviously, some processes must be overridden to acquire any language: a lexicon cannot be built of just the least marked segments, /ʊ/ and /ʌ/. A child learning English must acquire the abilities or keys that turn off processes (1) through (4) – first, in perception, as he learns to attend to the difference between words like hat and hot, even though he might at first still pronounce them alike, and eventually, in production, as he learns to pronounce them.
differently. A child learning Japanese must, of course, limit or turn off some substitutions, but he need not turn off (1)–(4), and so does not acquire the same abilities.

But at the same time, the speaker may need to alter the optimal or 'possible' segments left by the fortitions, when these segments are combined in real-time utterances, and he will thus allow certain lenitive substitutions to apply. In a language with no /u/, a speaker may produce [o] when /a/ appears adjacent to labial consonants, as in (5):

(5) Back vowels between labials must be labial.

A context-sensitive lenition like this does not alter the set of 'possible' segments – the phoneme inventory – because the existence of a phonetic motivation for the substitution in a particular context makes it possible for both speaker and hearer to discount or ignore the effects of the substitution.

The possible outcomes of the conflicting effects of the fortition (1) and the lenition (5) are illustrated in the chart below. In conservative United States dialects, neither applies, as in A, below. In the innovating dialects I know of, only (1) applies, as in B, or both (1) and (5) apply, as in C. But one would not be surprised to find a dialect like D, where (1) does not apply, but (5) does. Donegan and Stampe (1979) established, on the basis of optional fortitions and lenitions, that all fortitive processes must apply before all lenitive ones: fortitions first, lenitions last.

<table>
<thead>
<tr>
<th>Dialect:</th>
<th>cot</th>
<th>caught</th>
<th>bobble</th>
<th>bauble</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. (1) does not apply:</td>
<td>/kot/</td>
<td>/kot/</td>
<td>/baba/</td>
<td>/baba/</td>
</tr>
<tr>
<td>(5) does not apply:</td>
<td>/kat/</td>
<td>/kat/</td>
<td>/baba/</td>
<td>/baba/</td>
</tr>
<tr>
<td>B. (1) applies:</td>
<td>/kat/</td>
<td>/kat/</td>
<td>/baba/</td>
<td>/baba/</td>
</tr>
<tr>
<td>(5) applies:</td>
<td>/kat/</td>
<td>/kat/</td>
<td>/baba/</td>
<td>/baba/</td>
</tr>
<tr>
<td>C. (1) applies:</td>
<td>/kat/</td>
<td>/kat/</td>
<td>/baba/</td>
<td>/baba/</td>
</tr>
<tr>
<td>(5) does not apply:</td>
<td>/kat/</td>
<td>/kat/</td>
<td>/baba/</td>
<td>/baba/</td>
</tr>
<tr>
<td>D. (1) does not apply:</td>
<td>/kat/</td>
<td>/kat/</td>
<td>/baba/</td>
<td>/baba/</td>
</tr>
<tr>
<td>(5) applies:</td>
<td>/kat/</td>
<td>/kat/</td>
<td>/baba/</td>
<td>/baba/</td>
</tr>
</tbody>
</table>

The forms in /'s represent forms that are perceived and stored in long-term memory; those in ['s] represent forms as spoken.

3.2 The interaction of fortitions and lenitions creates phonemic representations

Whenever they can, speakers hear and remember 'impossible' sounds – sounds that are eliminated from the phonemic inventory
by fortitive constraints – as 'possible' sounds, so in dialects where (1) applies, only /a/ appears in phonemic forms. (Similarly, English speakers perceive and remember sounds like [e] in words like bend [bɛnd] as /ɛ/, or [i] in words like eighth [ɛɪθ] as /ɪ/) If possible, speakers do this by undoing the effects of lenitive processes to arrive at representations which the fortitive constraints do not rule out. The phoneme inventory of a language is thus the result of the interaction of the typically context-free fortitive processes and the context-sensitive lenitive processes. And the phonemic form of a word is the result of undoing subconsciously, in perception, enough lenitive processes to arrive at a form that is not ruled out by the obligatory fortitions. This is why speakers and hearers are typically unconscious of allophony.

Learning the phonemic system of a language, then, involves only learning which processes are not allowed to apply. A learner of dialect A need only realize that [a] sounds different from [α] in a single word, like caught, where its presence is unattributable to assimilation, to overcome process (1) – unrounding – in perception. On doing so, he can begin to lexicalize low labial vowels differently from low non-labials, and he marks (1) for eventual mastery in production. (Overcoming a process perceptually is of course necessary for overcoming it articulatorily, but it is not sufficient, which is why learners' perceptions and lexical representations are often more advanced than their productions.)

Learners of dialects B and C do not recognize the (a)/[α] difference, even in perception, and have not learned to pronounce [a] as an intentional, or permissible, sound. Thus they lexicalize all low non-palatal vowels as /a/.

But the B speakers, though subject to (1), can and do pronounce [α], by assimilation (5), in words like Bob, Mom, bomb, bobble, etc.10 These [α]'s are not perceived or remembered as different from the vowels of Don and cot. The B speakers apparently discount the labiality of these [α]'s as an effect of the labials in the environment. In fact, such speakers may be quite unconscious of any difference between their vowels in dog [dɑɡ] and Bob [bɑb], although A speakers, who maintain the original distinction, may notice the 'wrong' vowels in these words.

Learners who hear words like bobble, with [a] in potentially labializing environments, mark process (5) – labiality assimilation – for eventual mastery (suppression) in production. Speakers of dialect C seem to overcome this process.

Speakers of dialect A also overcome (5), and thus they eliminate (5) as a way of accounting for [α] in labial environments. The recognition that (5) does not apply in bobble means
that (5) does not account for the occurrence of the [o] in bauble. So when words like bauble [bobal] are encountered, the [o]'s in such words are perceived and remembered as /b/’s.

Learners of dialect D overcome (1) and thus distinguish between /a/ and /b/, but they do not need to overcome (5). Thus (5) remains as a way of accounting for the labiality of the [o] in either bauble or bobble. But since (1) has been overcome and /b/ is, consequently, a ‘possible’ sound – a phoneme – there is no need to undo (5) in the perception or lexicalization of either of these words; both [o]'s can represent /b/’s. (See Donegan 1985 for further discussion.)

3.3 Morphophonemics: undoing substitutions to make morphological connections

When the undoing of further processes in perception will also allow the expression of a morphological relationship, this too occurs, if, but only if, the speaker makes the morphological connection. To take a very simple example, handbag, usually pronounced ['hæm,baɡ], can be perceived phonemically (and possibly lexicalized) as '/hæm,bæɡ/ by undoing only the lenition which nasalizes vowels, since /hæm,bæɡ/ is a sequence of possible sounds (phonemes) in English. But ['hæm,bæɡ] can also be perceived as /'hænd,bæɡ/ (by undoing the lenition which assimilates a sequence of alveolar stops to a following labial, and that which degeminates stops), and it will be thus perceived if the speaker recognizes that it represents the morphemes hand and bag. Both /'hæm,bæɡ/ and /'hænd,bæɡ/ are possible phonological representations of handbag. /'hæm,bæɡ/, the most superficial possible phonological representation, is the phonemic representation – and thus we recognize its homophony with ham bag. /'hænd,bæɡ/, also a possible sequence of phonemes is, further, a morphophonemic representation. (See Stampe 1987 for further discussion.)

The term ‘morphophonemic’ is used here to refer to representations that are accessible through the undoing of phonological processes past the point where an acceptable phonemic representation is achieved. Arriving at a morphophonemic representation requires the undoing of one or more additional derivational steps in order to reveal a relationship with another form of the same morpheme. The term may also refer to the effect of a process which makes a different phonemic interpretation possible – e.g., the alveolar assimilation process that affects /'hænd,bæɡ/ has morphophonemic effects when nd → mb, where /m/ and /b/ are phonemes. This is not the same as ‘morphophonological’, a term
which refers to rules that appear to be partly phonological but are morphologically conditioned – or to representations that are not accessible through the undoing of purely phonological processes (see section 2.1).

4 Sound change and learnability

In the conventionalist view, the optimal grammar is the optimal combination of lexical forms and phonological rules. Because the phonetic forms of a language change over time, the optimal grammar to be derived from the phonetic forms will change. Conventionalist phonology appears to assume that segment inventories and phonological forms are acquired through the performance, by learners, of distributional analysis of remembered phonetic forms and through comparative evaluation of alternative grammars or grammar-and-lexicon combinations. (For a discussion of the improbabilities of this assumption, see Donegan 1985.) But there is an alternative explanation of how a child, who hears only phonetic forms, can arrive at adult-like phonemic representations as well as adult-like phonetic forms. The phonemes and the phonological forms of a language are not the results of grammatical analysis, but of automatic phonological processing.

In this theory of acquisition advocated here, the phonemic system of a language arises within the learner as a way of categorizing the elements of speech for storage in long-term memory. No distributional analysis, and no direct comparison of unrelated forms like bobble and bauble or cot and caught are required. Lexical representations are achieved and revised one word at a time, with reference to what one has learned about which processes apply and which do not. The learner goes from remembering less about each word to remembering more, from less mastery of perception and production and more reliance on 'defaults' to greater mastery, and from perceiving differences to producing them.

5 Changes in phonological representations

Phonological change begins with changes in pronunciation. Because the speaker perceives in terms specified by the phonological processes that apply in his speech, changes in learners' phonemic inventories or phonemic identities result 'automatically' from changes in the application of phonological processes in the speakers around them. Let us examine some
examples of how phonetic changes — changes in the application of phonological processes — result in changes in speakers’ phonological representations.

5.1 Loss of phonemes — merger of lexical classes
Perhaps the simplest example of a change in representation by the application of a phonological process is that of merger. Examples include the merger resulting from the unrounding of labiopalatal vowels, which took place not only in English (Brunner 1965), but also in Yiddish (Sapir 1915), and the German dialects of Darmstadt and Upper Austrian (Keller 1961). In these cases, a fortitive process,

(6) Palatal vowels must be non-labial

was allowed to apply, with the result that /y/ and /ø/ were pronounced [i] and [ɛ]. As the application became more widespread and general, learners stopped hearing [y] and [ø], and thus did not find it necessary to produce them, so /y/ was eliminated in favour of /i/, and /ø/ was eliminated in favour of /ɛ/. The fortition became a constraint on the inventory.

Similarly, in the United States dialects where caught and cot, dawn and Don, auto and Otto are homophonous, the application of (1) is responsible for the merger of /ɒ/ with /ɔ/. This process or constraint, applying in all styles and at all tempos, means that the speaker does not learn to distinguish [α] from [ɔ] at all.

At first, such a substitution may be introduced as an option. The innovating speakers, in that case, maintain the lexical distinction, with only an optional surface neutralization. But if the process applies to substitute [a] for [ɔ] with sufficient frequency, a subsequent generation of speakers may find it unnecessary to inhibit the process by mastering the pronunciation of [ɔ]. For them, the unrounding is obligatory, and there is no phoneme /ɔ/, but only /ɑ/, because when a fortition is obligatory for a speaker, it affects his lexical representations, ruling out the input as ‘impossible’. Such fortitive processes may constrain perception as well as pronunciation, so that speakers do not even perceive the distinction in the speech of those who retain it.

As noted in section 3.2, while the context-free fortition acts as a constraint on the phoneme inventory, eliminating /ɔ/ in favor of /ɑ/, a context-sensitive labiality assimilation can account for the phonetic instances of [ɔ], as in [mɒm], [ppɔp], [bɒb] for mom, pop, Bob, if these occur. But it can only do so as long as the learner does not notice forms like invariant [bɒm] bomb, or
[babal] bobble, which would make the labiality assimilation untenable.

5.1.1 The Labov-Yaeger-Steiner paradox
Labov et al. (LYS) (1972) reported on a rather puzzling situation which they encountered with a speaker (in central Pennsylvania) who appeared to merge [a] and [b] in reading or judging minimal pairs. They found that this speaker maintained a firm distinction between words with /a/ and words with /b/ – like cot and caught – in casual or connected speech, but lost the distinction in reading minimal pairs, and denied its existence when asked about the pairs directly. LYS pursued this apparent paradox of 'reported mergers' and found other instances where speakers distinguished in connected or casual speech forms which they merged in careful or closely monitored speech – and which they judged to be 'the same'.

Surprise at this state of affairs results from a number of assumptions: (a) We assume that if speakers produce a consistent difference between minimal pairs in casual speech the pairs must be represented as different in long-term memory. That means the words have different phonemic representations. However, it is often assumed (b) that speakers' careful pronunciations, as in reading word-lists, are more like their lexical representations than pronunciations drawn from connected speech. And (c), we assume that speakers make same/different judgements on the basis of their lexical representations, rather than on the basis of connected-speech forms. Accepting (a), that if a speaker makes a consistent distinction in connected speech, he maintains the distinction in the lexicon, let us examine assumptions (b) and (c).

Many optional phonological substitutions (alveolar place-assimilation, flapping, flap deletion, 'monophthongization' of [an] to [n], etc.) are characteristic of casual speech. It is our attention to substitutions of this kind that makes it appear (b) is true: that careful speech forms, to which these substitutions do not apply, are closer to speakers' lexical forms. But there are some optional phonological substitutions – fortitions – that apply only in slow or careful or exaggerated speech, as when prayed is overpronounced as [prə'rej] (cf. Donegan and Stampe 1979). If such optional substitution neutralizes a phonemic distinction, we may find that speakers actually merge, in careful speech, forms that they distinguish in connected speech. E.g., prayed becomes homophonous with parade. It appears that LYS's Pennsylvania speaker applied the low-vowel unrounding fortition only optionally, in careful speech. He produced both a low back-rounded
vowel /o/ and a low back-unrounded vowel /o/ in casual speech, but only the unrounded /o/ in careful or reflecting speech. The careful speech forms [hak] and [kot] of /o/-class words like hawk and caught are thus ‘farther from’ the lexical forms, than the casual forms [hok] and [kot] would be. Therefore (b) is not always true.

Assumption (c) is that speakers make same/different judgements on the basis of lexical representations. Speakers often do make such judgements on the basis of lexical forms when an optional merger affects connected speech. For example, latter and ladder may be judged as different even by speakers who pronounce both as [læDr] in connected speech. This occurs because speakers ordinarily base their judgements on careful pronunciations, to which the merging process does not apply. But in the case of an optional neutralization, there are two possible phonological representations for the neutralized form: the lexical representation is the representation before the substitution applies, but the output of the substitution is interpretable as a different phonemic form. Thus, speakers may also recognize that latter and ladder are pronounced the same, [læDr], by noticing that latter can be pronounced ‘with a d instead of a t’. When the merger applies only in careful speech, the sameness judgement is particularly apt to reflect the merger, rather than the different lexical forms. So although caught is [kot] in connected speech for LYS’s speaker, it is [kot] – just like cot – in careful speech. Unfortunately, LYS did not report on the Pennsylvania speaker’s ability to perceive the cot/caught distinction in the speech of others. We would expect that this speaker could perceive it.

In the section in which this paradox is presented, many of LYS’s informants appeared to be considering their careful speech in same/different judgements, or to be marginally aware of dual possible pronunciations (which suggest dual phonemic interpretations). Unfortunately, I cannot account here for all of the puzzles LYS presented. My intention is simply to point out that optional careful-speech fortitions and resultant dual phonemic perceptions must both be considered when evaluating and accounting for speakers’ judgements.

5.2 Changes in the phonemic forms of words
The application of a process that is lenitive and context-sensitive but that does not create alternations may, of course, change the phonemic representations of some words without changing the phoneme inventory. In many dialects of English where /o/ and /o/
both exist, /l/ has become [o] after [w] (and before a
tautosyllabic liquid), as in wall, squall, war, quart, [wol, skwol, 
wor, kwort]. Learners encountering invariant [o] in [wol] might 
attribute it to an assimilation if [o] were ruled out of the 
phoneme inventory by a fortition. But since, in most dialects, this 
fortition is suppressed to allow words like caught [kɔt], and since 
these forms involve no variation or alternations with /l/, there is 
no reason for language learners to interpret the [o] here as 
anything other than /l/.

There are, however, more interesting cases of changes in the 
phonemic representation of a class of words because of the 
addition of a lenition process. The case of ‘intrusive r’ in English 
seems to be such a case. It is well known that ‘r-insertion’ occurs 
only in dialects in which syllable-final r is de-rhotacized. In RP, 
for example,

(7) r in a syllable-final (before consonant or pause) loses its 
r-colouring, becoming ə.

De-rhotacizing [r] leaves a non-syllabic schwa, which may 
coalesce with a preceding syllabic schwa. Thus, we find final [ə] 
or [ə] phrase finally or before consonants, as in

A. 

butter [bʌtə]  favour [feɪvə]
hear [hɪə]  pour [pɔːr].

But pre-vocally, where [r] can begin the syllable, [r] appears:

B. 

butter it [bʌtəɾ ɪt]  favour it [feɪvəɾ ɪt]
hear it [hɪəɾ ɪt],  pour it [pɔɾ ɪt].

In RP, and other dialects which have ‘intrusive r’, words that 
have final schwa when pronounced alone or before consonants, 
e.g.

C. 

India ['ɪndiə]  idea [ai'diə]  drama ['draːmə],

are pronounced with ‘linking’ r when they occur before words 
that begin with a vowel:

D. 

India and Pakistan ['ɪndiə ən 'pækiʃən],
drama and music ['draːmə ən 'mjʊzɪk]
idea of ['aɪdiə əv].

This innovative [r] may also occur – though ‘less frequently’ after
118  ON THE PHONETIC BASIS OF PHONOLOGICAL CHANGE

final /əː/ or /ɔː/, as in

E.

Shah of Persia  ['ʃaːr əv 'paːrə]

law and order  ['laːr ənd 'ɔ:dər]  (Gimson 1962: 203ff.).

Gimson describes this ‘intrusive r’ as occurring by analogy. In generative phonology, the insertion of r in these dialects has been described as a ‘rule inversion’ (Vennemann 1972: 216). This means that speakers of dialects which have an ‘r-deletion’ rule (r in syllable-offsets is de-rhotacized) reinterpret the resulting data and create a new rule, one which inserts r after certain word-final vowels. (It would be possible to say the insertion occurs after word-final schwa; insertion after final /əː/ or /ɔː/ may depend on the pronunciation of these vowels with centering glides, as [əə] or [ɔɔ].) But neither ‘analogy’ nor the creation of an inverted and perhaps more general rule can explain the difficulty of pronunciations without the intrusive r. Gimson says that pronunciations like I saw it /əI 'sɔːr ɪ/ or drawing /'drɔːrɪŋ/, are ‘generally disapproved of’, and that ‘it is likely that many RP speakers have to make a conscious effort to avoid the use of such forms’. He says, in fact, that they may have to insert a glottal stop or pause to do so (1962: 204). Attempts not to analogize or generalize would not create such a pronunciation difficulty for the speaker.

So instead, we might take the following view.13 In these dialects, the lenition, (7), applies obligatorily. Therefore speakers find final r’s difficult or impossible to pronounce, and they produce alternations like those between the A and B forms above. Learners can account for such alternations by ‘undoing’ the lenition (7), and they thus arrive at underlying forms with final /r/. Words like hear and favour, with a final [ə] or [ɔ] are interpreted as having final /r/ to account for the B forms. The A forms are pronounced before pause or consonant because of (7), and the B forms occur before a vowel, since the r can be syllabified with the following vowel.14

But learners may also undo the r-to-schwa lenition in perceiving words like India and drama, which have final schwas that did not originate in r’s: thus, /ɪndɪər/, /ˈdramaːr/. And if words like draw and Shah – and even baa – are pronounced with schwa offglides [drəə], [ʃaː], [bæə] they too can be interpreted as having final vowel-plus-/r/: as /drɔr/, /ʃa r/, /bæə/. But then, before vowels, where de-rhotacization does not apply, these /r/’s will actually be pronounced. (This may happen even though the words are often pronounced [drɔː]; [ʃaː], etc., since these may be perceived as assimilated forms of [drəə], [ʃəə], which are in turn
perceived as de-rhotacized forms of /drər/, /ʃər/.

The 'intrusive r' does not, then, intrude because the speaker makes up an r-insertion rule. Instead, the r appears by analysis, when speakers assume that, because some final schwas represent /r/’s, other final schwas do so as well. Assuming that final schwas are /r/’s allows certain vowels to be eliminated from the phoneme inventory: the long vowels /ɔː/, /ʌː/, and /ɜː/ (just those that cannot be analyzed as vowel plus /ʃ/ or /u/) would be represented as /ɔə/, /ʌə/, and /ɜə/; and the diphthongs described as /ɪə/, /ɛə/, /ɔə/, /ʊə/ would be /ɪr/, /ɛr/, /ɔr/, /ʊr/. Lexical postvocalic /r/’s are required in any case, but having /r/ rather than length or /ə/ means that fortitive constraints can be maintained which eliminate distinctive length or offglides other than /ʃ/ or /u/. Speakers without intrusive r must overcome these constraints to acquire long vowels and final /a/’s.

Thus Shah, draw, baa, and India are hard to pronounce without r, in linking environments like Shah of, draw it, because there is, in these environments, no reason to de-rhotacize their final /r/’s. The change is a matter of perception: speakers with intrusive r’s perceive final [ə]’s as /r/’s, by attributing them to the application of (7). Speakers without intrusive r’s can perceive final schwas as /ə/’s (or they can ignore them, as the results of a lenition that centralizes the offsets of long low vowels) in words where alternations do not reveal the presence of /r/’s.15

5.3 Addition of phonemes to the inventory, process ‘loss’
New applications of lenitive substitutions may not only cause changes in the phonological forms of words; they may also result in changes in the phoneme inventory. Further, the innovative application of one lenition may, in obscuring the environment for another lenition, cause that ‘earlier’ lenition to cease to apply.

Old High German (OHG) umlaut is a well-known but instructive example of both these results (cf. Stampe 1987: 294). This context-sensitive palatality assimilation (a lenition) might be stated:

(8) A vowel is palatalized if a high palatal vowel follows in the same metrical foot.

This created OHG forms like betti ‘bed’ and stein ‘stone’ (cf. Gothic badi, stains). Assimilated (short or long) /u/ and /o/ became [y] and [e], as /a/ became [e]. For example, in OHG, muːs ‘mouse’ alternated with mʊːsi ‘mice’, fuɔt ‘foot’ with fiʊsɪ ‘feet’, and gast ‘guest’ with gɛstɪ ‘guests’. Twaddell (1938) points out that the [y] and [e] outputs of umlaut were indicated in
spelling only in Middle High German, after the palatal environment which occasioned the umlaut had been eliminated. But the [e] result of umlaut was spelled in OHG. Since OHG had an /e/ phoneme, it was possible for speakers to perceive the [e] result of umlaut as the phoneme /e/. (Speakers did not need to undo the umlaut lenition, except to connect alternants.) And, as the spellings indicate, the OHG scribes did perceive this umlaut product, [e], as /e/.

OHG had no /y/ or /ø/ phonemes, however. We can say that [y] and [ø] were ruled out of the phonemic inventory by fortitive constraints against labiopalatal vowels:

(9) **Labial vowels must be non-palatal.**
(10) **Palatal vowels must be non-labial.** See (6) on p. 114.

In perceiving [y] and [ø], speakers apparently undid the umlaut process which created them; in effect, they discounted their palatal quality as an effect of the following palatal vowel, and perceived and spelled [y] and [ø] as /u/ and /ø/. Only when a later change,

(11) **Unstressed vowels must be non-palatal.**

reduced the unaccented palatal vowels which occasioned the assimilation did speakers begin to perceive and spell [y] and [ø] as separate phonemes.

Such re-phonemicization must take place in the learners of the language. During a time when the vowel reduction was variable or optional, learners encountered phonetic forms like [mus] 'mouse', [mysi] ~ [myisa] 'mice'. They could interpret these as /muses/ and /my:si/ by undoing both the vowel reduction (11), and the palatality assimilation (8). But at some point, the reduction (11) became sufficiently exceptionless to make the unstressed palatal vowel impossible for learners to perceive. Learners encountering only phonetic forms like [mus] and [my:sa] no longer had reason to interpret the plural suffix as other than /ø/. Therefore, they could no longer discount the palatal quality of [y] as the result of an assimilation to a final /i/. At that point, learners had to forgo the fortitive processes (9) and (10) that had constrained the phoneme inventory, and learn to produce, perceive, and remember labiopalatal vowels. They were, in other words, forced to admit /y/ and /ø/ as phonemes. Thus, the obligatory vowel reduction process (11), by completely obscuring the motivation for the umlaut assimilation (8) resulted in speakers' having to overcome the fortitions that had eliminated labiopalatal vowels.
At this point, the vowel pairings created by umlaut became usable in morphological marking, and they eventually acquired significance in marking plurals, subjunctives, diminutives, etc. Even when the fortition represented by (10) or (6) re-asserted itself in some German dialects, unrounding the labiopalatals as described in section 5.1, the pairings, though entirely arbitrary from a synchronic point of view, retained their morphological significance (cf. Sapir 1921, ch. 8).

5.4 Phonemic split and complementary distribution
When changes in the application of lenitions make a fortitive constraint untenable, as in the MHG case, the phoneme inventory changes. And when a fortition applies obligatorily, as in American dialects where /ɔ/ > /œ/, the fortition changes the phoneme inventory. It is change in the application of fortitions that allows (or creates) the change in the inventory.

The fortitive constraints in these two examples, labial vowels are non-palatal and low vowels are non-labial, are context-free; this is typical of fortitions. But a fortitive change may apply context-sensitively, applying only in stressed syllables, in lengthening contexts, at intonation peaks, in dissimilative environments, etc. For example, in most British and American dialects, early Modern English [əi] became [œi] and [ʌi] became [ou] - that is, the following fortitions applied:

\[
\begin{align*}
(12) & \text{ A non-palatal syllabic that precedes a palatal glide must be low.} \\
(13) & \text{ A non-labial syllabic that precedes a labial glide must be low.}
\end{align*}
\]

If an obligatory fortition constrains phonemic representations, a fortition that becomes obligatory will change speakers' phonemic representations as (12) and (13) did. This may happen even when the fortition applies only in specific environments.

In certain United States and Canadian dialects, the vowel corresponding to Middle English long iː is represented in some contexts by [œi] and in others by [ʌi]. The distribution is often described, somewhat over-simply, as [ʌi] before voiceless consonants and [œi] elsewhere. The rule has been called 'Canadian Raising' (Chambers 1973), but it is not limited to Canadian dialects, nor is it necessarily a raising - no evidence has been offered that [œi] is basic and [ʌi] is derived. I follow Stampe (1972) and Gregg (1973) in suggesting that the United States and Canadian dialects which retain [ʌi] have not quite completed the Great Vowel Shift change of ME iː to /œi/, and that 'Canadian
Raising' is actually the result of lowering the \([\text{Ai}]\) reflex of ME i: to \([\text{ai}]\) only in certain environments, rather than generally.17 Gregg adduces evidence for this view from Anglo-Irish, in which ME i: remains /ai/, and from Scottish and Scotch-Irish, in which \([\text{Ai}]\) and \([\text{ai}]\) apparently diverged early, and underwent a phonemic split.

This alternation (or split) is clearly related to the process by which vowels are lengthened in open syllables or before voiced (lenis) consonants. Long vowels are, ceteris paribus, more susceptible to lowering than their short counterparts (Donegan 1978), and basically, /\text{ai}/ is lowered to [\text{ai}] where it is lengthened, and it remains [\text{Ai}] where it is short. The conditions that create sufficient lengthening to condition the quality change apparently differ from dialect to dialect. They include not only the voicing and continuance of the following consonant, but also the number and accentuation of following syllables within the accent-group.

For speakers who have no quality difference in words with /\text{ai}/, we expect a length difference between the vowels of write [\text{rait}] and ride [\text{rai:d}], as between those of kit [\text{kIt}] and kid [\text{kI:d}]. Thus fly, ride, tide, gibe, drive, tithe (with final [\text{ð}]), rise, etc. have long vowels, while write, tight, ripe, life, blithe (with final [\text{θ}]), rice, etc. have short. In the North American dialects with an [\text{Ai}]-[\text{ai}] alternation, the words in the first set have [\text{ai}], and those in the second have [\text{Ai}]..18

For some Michigan, Minnesota, and New York speakers, the vowel in words like cider, spider, and idle, as well as in miter, title, is [\text{Ai}]. In these disyllables, the nucleus is mid, regardless of the voicing of the following consonant, because the presence of a following unstressed syllable within the stress-group, or foot, shortens the first syllable (cf. Lehiste 1971a, b). The stressed nucleus is thus not lengthened in words like spider and pilot (as it would be in spied or pile). Because their phonetic shapes do not provide the conditions for lengthening, these disyllables have not undergone the length-dependent lowering.

Note that these words, though disyllabic, are monomorphemic. In words (or phrases) of similar shape but bimorphemic structure, like glider, slider, tidal, pile it, the vowel of the stressed syllable in these dialects is [\text{ai}]. But why should spider, idle and pilot have [\text{Ai}], while slider, tidal and pile it have [\text{ai}]? The obvious answer is that the two-morpheme forms with [\text{ai}] maintain the same vowel that their one-syllable bases have. That is, [\text{ai}] occurs in the derived forms, slider, tidal, pile it because it occurs in the one-syllable base forms, slide, tide, pile. This might be expressed formally by having a morpheme boundary condition
the stressed-vowel change.

But this response leaves us with a further problem: the \[^{Ai}\] \[^{oij}\] alternation is usually regarded as allophonic, and allophonic alternations or changes do not ordinarily depend on morpheme identity. And we would not expect them to. If, as has been repeatedly observed (by Sapir 1921, Bloomfield 1933, Swadesh 1934, and others), speakers are not aware of allophony, we would expect them to make allophonic changes under conditions that are purely phonetic (and prosodic) – not morphological. But the lowering of \[^{ai}\] in these disyllables does seem to depend on morphemic identity.

The explanation seems to be that, in becoming obligatory, the lowering, a fortition, changed the phonemic forms of the words in which it applied. When the process,

\[(12') A \text{ long non-palatal syllabic that precedes a palatal glide must be low.}\]

was optional (as it may have been at some point in history), then /\[^{rAijd}\]/ could be pronounced \[^{rajd}\] or \[^{rajd}\]. But when (12') became obligatory, then

/\[^{rajd}\]/, always pronounced \[^{rajd}\], became /\[^{rajdl}\]/ \[^{ride}\], while

/\[^{rajt}\]/, pronounced \[^{rajt}\], remained /\[^{rajt}\]/ \[^{write}\].

With affixation and flapping,

/\[^{rajt}\] \[\&/r\] becomes \[^{rajDr}\], and

/\[^{rajd}\] \[\&/r\] becomes \[^{rajDr}\].

The lowered vowel, being phonemic, now no longer depends on a particular phonetic environment. The evidence for the phonemic nature of the change is that the words with lowered nuclei retain these low nuclei even when an affix makes them disyllabic and shortens the stressed vowel. It is notable that forms like pilot and pile it may constitute minimal pairs in such dialects. It is also worth noticing that 'Along came a sp[^{ai}]der' does not rhyme with 'and sat down bes[^{oi}]de her' for these speakers.

Vance (1987) argues that the difference between \[^{ai}\] and \[^{Ai}\] is phonemic, at least in some dialects. He notes the ease with which 'naive' informants who use \[^{ai}\] and \[^{Ai}\] are able to attend to this difference and make judgements about it. This ease is contrasted with the difficulty of training ordinary speakers to detect allophonic differences like that between front vs. back [k] and clear vs. dark [l]. Vance also points to the irregularities in the distribution of \[^{ai}\] and \[^{Ai}\], and to some minimal pairs (like idle/idol). He notes that the difference between an \[^{ai}\]/-\[^{Ai}\]/
distinction and the presence of /oɪ/ in all environments seems to involve lexical diffusion, and that changes that occur by lexical diffusion are always phonemic (cf. Chen and Wang 1975, Krishnamurti 1978).

Once a speaker has determined that two sounds, like [ɔɪ] and [ʌɪ], are different phonemes and must be represented differently in long-term memory, his representations of individual words that contain either sound will depend on the pronunciations of the speakers from whom he learns each word. It does not matter whether the other speaker makes a distinction or not; if the new word is heard with [ɔɪ], the learner will remember /ɔɪ/, and if the new word is heard with [ʌɪ], the learner will remember /ʌɪ/. This means that the pronunciations of speakers who make a distinction will, for the most part, agree with those of speakers for whom the alternation is automatic. It also accounts for the discrepancies and occasional indecisions among the speakers Vance questioned. (Hearing multiple variants may create indecision in the learner, although it will not necessarily do so.)

The distribution of [ɔɪ] and [ʌɪ] is somewhat different in Baltimore, Maryland, from the distribution in the dialects Vance discussed. In monosyllables and monomorphemic disyllables, the following consonant phoneme predicts the variant: [ʌɪ] occurs before voiceless consonants, and [ɔɪ] occurs before voiced consonants or finally: so tight, miter, and title have [ʌɪ], and tide, spider, idle, and tie have [ɔɪ]. Even in polysyllables, the difference seems to be predictable: before an unstressed syllable, the following consonant is predictive, as above: glycogen and nitrogen have [ʌɪ], but gyroscope, hibernate, and hydrogen have [ɔɪ]. Before a stressed syllable, the vowel is [ɔɪ]: hypotenuse, isosceles, vituperate, iconic (bicycle and bifocals, with secondary stress on the second syllable, fall into this category), although here there are some exceptions: nitrate, python, have [ʌɪ] before a secondary stress, and psychotic, licentious, micrometer are at least possible with [ʌɪ] before primary stress, where [ɔɪ] is expected.

Yet the same ability to identify the difference that Vance noted as evidence of phonemic status also exists for Baltimore speakers. Although raider and rater, kiddie and kitty, with medial flaps, are homophones, writer and rider with identical medial flaps are not. In this dialect, as in Vance's, it seems that [ɔɪ] and [ʌɪ] are different phonemes, in spite of the fact that here they are in complementary or near-complementary distribution. This in turn suggests that an obligatory fortitive constraint, even if it is context sensitive, may affect the phoneme inventory. This
SUMMARY

supports the idea that it is the obligatory application of fortitions (along with the possibility of perceiving via lenitions) – rather than the distributional analysis of a mass of remembered phonetic forms – that creates the phoneme inventory. (Cf. the discussion in s. 4.)

6 Summary

The examples presented here were chosen to show that every phonological change – from a ‘low-level’ nuance of pronunciation to a radical restructuring of perception – can be understood as the failure by speakers to overcome a phonetic constraint that past speakers did overcome. Such a constraint, manifested as a substitution, was already a part of a dynamic mental system of phonetically motivated substitutions that constitute the phonology. It is this same robust and subtle system that, from the close of the babbling period in infancy, and throughout the life of the individual, provides a perceptually and/or articulatorily optimal substitute for any unpronounceable utterance. (Recall that this unpronounceability extends even to the phonemic representations of the native language – in English, can’t /kænt/ cannot be pronounced without aspiration or vowel nasalization, and in connected speech it is ordinarily [kʰæŋt].) Whether the unpronounceability is due to the intrinsic difficulty of the intended utterance, or its tempo, or its novelty (as in foreign words, or the results of tongue-slips or Pig-Latins), or whether it is due to the speaker’s condition (fatigue, nervousness, or the dentist’s hands in his mouth), the phonology instantly puts the appropriate substitute at the ‘tip of one’s tongue’.

The pioneers of phonology – Baudouin, Passy, Jespersen, Sapir, Meillet, Jakobson – recognized that all sound change (as opposed to analogical change) is simply a failure of new speakers to overcome some aspect of this inner system of ‘phonetic tendencies’. The problem of how sound changes become part of the phonology, and the problem of how to impose ‘naturalness constraints’ on phonological descriptions, are pseudo-problems that have arisen only because modern linguistics, obsessed with economical descriptions, has separated phonology from phonetics. But they are inseparable: the phonology is the phonetics, and phonological change is phonetic change.

Acknowledgement

Sincere thanks to Ann Peters, Ken Rehg, and David Stampe for all their help. Errors, of course, are my own.
Notes

1. For the sake of brevity, I omit ‘intermediate’ substitutions that involve only tense/lax differences (e.g. [ɔ] → [ŋ] → [o]) or diphthongizations limited to just the onset or the offset of the vowel (e.g. [ɔ] → [əŋ] → [oŋ], or [ɔ] → [ʊŋ] → [ʊŋ]).

2. Some learners easily acquire keys that elude others. For example, some children learning English get the vowels nearly right, making few substitutions, almost from the beginning, while others make do with a small set of substitutes for a rather long time: contrast Amahl Smith (Smith 1973) and Joan Velten (Velten 1943). Such differences may be idiosyncratic, or they may depend on developmental factors or on the learner’s prosodic system.

3. E.g., US speakers learn to distinguish latter [lætr] from ladder [lædr], but they do not do so in ordinary speech, and producing the distinction is perceived as requiring special effort.

4. The distinction goes back to Baudouin de Courtenay (1895), who called the morphologically conditioned alternations of morphophonology ‘correlations’, as opposed to the phonetically motivated ‘divergences’ of what is here called phonology. (The division between ‘post-lexical’ and ‘lexical’ rule applications is not presented as a difference between rules with and without synchronic phonetic motivation.)

5. Morphological rules are insensitive to prosodic factors like tempo and intonation, and to physical states like drunkenness or tiredness. They never affect or depend on non-distinctive features; rather, they always substitute one phoneme for another. See Donegan and Stampe (1979) for further characterization of the difference between such conventional rules and phonological processes.

6. Cf. Keating’s remarks that phonetics is being viewed, more and more, as ‘largely the same sort of creature as the phonology, that is, a formal system of rules’, and that ‘it is too early to decide the exact division of labour between phonological and phonetic rules’ or even ‘what is at stake in positing such a division’ (1988: 288, cf. 287-91). I am arguing that there is no such division.

7. In word pairs like divinity/divinity, wordy/wordiness, profound/profundity, etc., there is a relationship between ‘long’ and ‘short’ phonemes /oʊ, ʌ/ /oy, al/, and /u/, Al. But no evidence has ever been presented that this relationship is synchronically phonological and based on features, rather than morphological and based simply on conventional pairings of phonemes. If the change of a feature in one member of the pair implied anything about the pronunciation of the other member, one might make an argument for a feature description that unifies the pairs. For example, if the dialectal fronting of /ɔʊ/ to /ɒ/ in wordy and /ʌ/ to /æ/ in profound implied anything about the pronunciation of /a/ in wordy or /ɛ/ and /æ/ in profound, one might argue for a feature relationship between /oʊ/ and /a/, or between /au/ and /æ/ – but arguments for a synchronic phonetic basis have not stood up to scrutiny (see, for example, Manaster-Ramer 1981).
8. The admissibility theory of classical generative phonology at least made an attempt at such prediction. But, as Greg Lee has pointed out, it implied that an inadmissible word like Gdansk would be pronounced by English speakers as [stansk].

9. In Arabic, or Brunei Malay, with three vowels /i, a, u/ and /e/ are ruled out, as well as /æ/ and /o/. Thus, (1) and (2) apply to non-high vowels, and (3) and (4) result in high, instead of non-low, vowels.

10. This is like the pronunciation of nasalized vowels before nasals, by speakers who cannot pronounce such vowels 'intentionally' in non-nasal environments.

11. In English, the unrounding of mid vowels preceded that of high vowels, and the unrounding of short vowels preceded that of their long counterparts.

12. "Fall" indicates decreasing sonority. A syllable-fall includes the stressed syllable and extends to the end of the syllable. A measure-fall includes the stressed syllable and extends to the end of the measure, or foot (Donegan and Stampe 1978).

13. This is a proposal of David Stampe's.

14. Speakers of dialects like Southern United States or Hawaiian English, with no linking r's at all, need not have a difference in the application of (11i) or any other substitution. They appear, instead, to have a difference in the syllabification of r, so that r does not resyllabify across a word boundary, with the result that the r's do not appear before vowels:

\[
\begin{align*}
\text{hear} & [\text{h12}] & \text{hear it} & [\text{h13 it}] \\
\text{mother} & [\text{m10a}] & \text{mother is} & [\text{m10a 12}]
\end{align*}
\]

15. As Gimson notes (204), "Spelling consciousness remains an inhibiting factor" in this development. The syllable final r's that occur in unconventional spellings sometimes represent these 'intrusive' r's. Louisa May Alcott's 'Marmie' in the New England speech of her Little Women would have been ['m1ami], and A. A. Milne's 'Eeyore' in Winnie-the-Pooh is, of course, ['iːtɜːr] or ['iiːtɜːr], which is how the donkey says his own name.

16. This alternation is remarkably widespread. It occurs not only in Canada and certain northern states of the US, but also in Maryland, Virginia, the Carolinas, and Georgia. It seems to underlie the distinction between the [əi] of tight (< [aɪ]) and the [ae] of tide (< [æ] < [ai]) in much of the Southern United States. It is also characteristic of Hawaiian English.

17. Like Stampe, I regard the [əu]-[au] alternation, too, (as in cloud/cloud) as a result of lowering. Gregg notes only that ME ə > a in Scots, Scotch-Irish and Canadian is a 'parallel but somewhat different case' (1973: 144).

18. It has been claimed that the slight lengthening that occurs before voiced consonants even when unstressed syllables follow produces a quantity distinction between writer [roiDɪ] and rider [roiːDɪ] for some speakers. Such a lengthening would also distinguish kity [kiDi] and kiddie [kiDi]. I know of no studies that confirm either
ON THE PHONETIC BASIS OF PHONOLOGICAL CHANGE

distinction. The degree of lengthening in such disyllables is far less than in monosyllables, and may in fact be negligible (cf. Lisker 1975; Lehiste 1971b).

19. The news that I was running an experiment to test whether Ohio speakers could perceive the difference between words like rider and writer puzzled my Baltimorean family, even when I pointed out the sameness of the t and d pronunciations in these words. Although all the speakers I asked agreed that kitty and kiddie were homophones, the idea that rider/writer were similarly indistinguishable, in Ohio, was met with disbelief all around. On the other hand, the Ohio speakers thought that I was setting them a completely impossible discrimination task.

20. The status of [ai] and [aj] in these dialects has been briefly indicated here. It merits further investigation because it calls into question the distributional criteria by which phonemes have always been established (though these criteria have never been altogether satisfactory).

References


REFERENCES


ON THE PHONETIC BASIS OF PHONOLOGICAL CHANGE


LISLER, LEIGH (1957) 'Closure duration and the intervocalic voiced-
voiceless distinction in English', Language 33: 42-9.


SAPIR, E. (1915) 'Notes on Judeo-German phonology', The Jewish Quarterly Review, n.s. 6: 231-66. (Reprinted in Mandelbaum 1949, pp. 206-12.)


